

Neuronale Denkfehler

Künstliche Intelligenz: zu naiv, um schlau zu sein



Fehlerbehaftete KI	Seite 128
Praxis mit Tücken	Seite 130
Konstruktionsfehler neuronaler Netze	Seite 134

Im Wettkampf Mensch gegen KI gibt der Homo Sapiens fast täglich eine klägliche Figur ab. Doch die meisten Erfolgs-Stories sind oberflächlich und verschweigen, wo es überall noch klemmt. Tatsächlich bleiben jede Menge Fragen offen, auf die die heute verfügbare Technik noch keine Antworten hat.



Bild: Fair Blog

Fehler in der Face-Swap-KI: Punkt, Punkt, Komma, Strich – fertig ist das Herdgesicht.

Von Andrea Trinkwalder

Noch nie schien der Traum von der Technik, die dem Menschen im Denken ebenbürtig oder gar überlegen ist, so nahe: Algorithmen erkennen nicht nur Objekte auf Fotos. Sie fungieren auch als persönliche Assistenten, schlagen den Weltmeister in Strategiespielen, die für normal begabte Menschen viel zu komplex sind, und komponieren, malen oder schreiben Geschichten.

Häufig brauchen sie dafür nicht mal mehr menschliche Vorbilder. Die neueste AlphaGo-Version der Google-Tochter DeepMind hat das asiatische Brettspiel Go nicht wie bisher aus der Beobachtung

menschlicher Matches gelernt, sondern von Grund auf, indem sie mit einem Klon ihrer selbst trainierte.

Künstliche Naivität

Doch nicht jedes harte Training steigert die Intelligenz. Künstlichen neuronalen Netzen unterlaufen bisweilen ausgesprochen dümmlich wirkende Fehler, die der menschliche Betrachter schlicht nicht nachvollziehen kann. Viele Fehler sind lustig. Sie füllen ganze Blogs, etwa der geniale Kunstgriff einer Face-Swap-KI, die das Gesicht eines Babys mit den Drehgriffen eines Herds vertauschte, anstatt mit dem Kopf des Vaters. Andere haben schwerwiegende Folgen, beispielsweise KI-Systeme im Bewerber-Management

oder autonome Fahrzeuge, die eine Verkehrssituation falsch einschätzen.

Denn während Wissenschaftler noch intensiv nach Lösungen suchen, die die Entscheidungen maschinell trainierter Systeme transparenter machen, laufen seit Jahren intensive Feldversuche, wie ein Blick in die USA zeigt. Dort geben Algorithmen Sozialprognosen über Straftäter ab und sortieren Stellenbewerber aus – häufig zum Nachteil sozial schwächerer Gruppen. Kritiker fürchten deshalb, dass hart erkämpfte Errungenschaften wie die Gleichstellung von Menschen unabhängig von Geschlecht und Herkunft durch die vermeintlich neutrale Technik peu à peu wieder aufgeweicht werden.

Dass künstliche neuronale Netze nicht immer die Ergebnisse liefern, die der Mensch erwartet, hat mehrere Gründe. Häufig liegt der Fehler in den Trainingsdaten, die dem Netz nicht helfen, genau die Fähigkeiten auszubilden, die es zum Lösen der gestellten Aufgabe braucht.

Aber es liegt auch an der Architektur selbst, die dem komplexen Zusammenspiel menschlicher Neuronen und Synapsen nicht annähernd ebenbürtig ist. Deshalb lernt der Mensch viel effizienter als die KI, denkt in logischen Zusammenhängen anstatt nur in Mustern und kann Humor und offensichtlichen Nonsens einordnen. Ein maschinell trainiertes Bilderkennungssystem hingegen lässt sich von kuriosen Szenen komplett verwirren: Montiert man etwa eine Giraffe in ein Wohnzimmerfoto, leidet die Genauigkeit der Objekterkennung.

Wie viele kleine und große Schwierigkeiten eine KI in der Praxis fehlerbehaftet oder gar unbrauchbar machen können, beschreibt der Artikel auf Seite 130. Warum sich mit den State-of-the-Art-Verfahren fundamentale Prinzipien menschlichen Denkens nicht simuliert lassen, erfahren Sie ab Seite 134. (atr@ct.de) **ct**

Trainingslager

Künstliche neuronale Netze (KNN) bestehen aus mehreren Schichten von Neuronen, die über Synapsen miteinander verbunden sind. Während das menschliche Original mit chemischen und elektrischen Reaktionen auf einen Reiz reagiert, arbeiten die künstlichen Neuronen mit einfachen mathematischen Funktionen, um Reize über die Synapsen an die Neuronen der nächsten Schicht zu verteilen.

Anhand kategorisierter Trainingsdaten – etwa Fotos oder einfacher Sätze – und einer Feedback-Funktion lernt das KNN in zahllosen Durchläufen, die Parameter seiner Basisfunktionen so zu justieren, dass es die Eingabe immer besser interpretieren kann. Zur Objekterkennung etwa bildet es Funktionen heraus, die Kanten und Muster erkennen und in tieferen Schichten zu charakteristischen Objektteilen kombinieren. Zur Spracherkennung bilden sich Filter, die Tonfrequenzen mit Buchstaben, Silben und Se-

quenzen aus ganzen Wörtern matchen. Dabei handelt es sich aber nicht um echte Abbildungen, sondern um Vorhersagen: Das System gibt eine Prognose ab, mit welcher Wahrscheinlichkeit etwa ein Mensch, ein Hund oder ein Auto auf einem Bild zu sehen ist.

Für das Training solcher Konstrukte benötigt man einige tausend kategorisierte Datensätze: also Datensätze, in denen die Eingabe und das gewünschte Ergebnis enthalten sind. Man spricht dann von überwachtem Lernen. Die Netze sind hochspezifisch und von den KI-Experten händisch auf den jeweiligen Einsatzzweck getrimmt. Ein neuronales Netz zur Bilderkennung lässt sich also nicht einfach auf Spracherkennung trainieren, indem man es mit Sprachproben füttert. Ziel der KI-Forscher ist es, unüberwachtes Lernen zu verbessern, um künftig die Abhängigkeit von kategorisierten Datensätzen zu reduzieren.